Trained-Once Device-Free Crowd Counting and Occupancy Estimation Using WiFi: A Doppler Spectrum Based Approach

Simone Di Domenico, Giovanni Pecoraro, Ernestina Cianca, Mauro De Sanctis
Department of Electronics Engineering,
University of Rome Tor Vergata,
Rome, Italy
Email: mauro.de.sanctis@uniroma2.it

***Abstract*—This paper presents a WiFi-based device-free crowd counting and occupancy estimation system that can be used in rooms/environments different from the ones in which the training process has been performed. Therefore, crowd counting is achieved without requiring another training phase in each new environment. The proposed approach analyzes the shape of the Doppler spectrum of the received signal which is correlated to the number of people moving in the monitored environment. Unlike a radar-like approach, the use of a reference signal is not required. Experimental results are presented for two different rooms/environments without any constraint on the movements of the volunteers.**

***Keywords*-Device-free, People Counting, WiFi, RF sensing, Crowd Counting, CSI, RSSI.**

I. Introduction

Crowd counting is the process of estimating the number of people in a given area, in either an open or closed environment, for both safety critical and uncritical situations. The tracking and monitoring of human queues or people sitting in a waiting room may be used for improving the offered services in areas like retail stores or airports, hospitals and theme parks [1], [2]. Such improved services result in a better user satisfaction and in reduced service costs.

Usually, the required accuracy of the estimation process decreases as the number of people increases. Therefore, if on one hand it is important to distinguish between empty room, one person in a room or two people in a room (i.e. crowd counting), usually for a higher number of people it is sufficient to provide a density estimation like stating what is the probability that there are, for instance, 3 to 5 people, or 6-10, or 11-20, and so on (i.e. crowd density estimation).

In this paper, we focus on RF-based device-free sensing which does not require any specific user cooperation [3]. RF-based device-free crowd counting or density estimation techniques are in general based on some kind of measurement on the RF propagation channel as a function of the presence of a variable number of people. Measurements performed at the receiver side include Received Signal Strength Indicator (RSSI), Channel State Information (CSI) and Channel Quality Indicator (CQI) [3], [4].

So far, most of the approaches for crowd counting are based on measuring and properly processing the RSSI of the RF signal [5], [6], [7], [8], [9], [10]. To describe the propagation channel variations, a more recent approach uses the CSI measurements.

It is worth noting that all these works require a training phase, which *must be performed in the same environment* as the one in which the crowd counting is performed, with the same setting conditions (i.e. same position of transmitters and receivers, same furniture, etc.). Therefore, these techniques cannot be used in environments different from the ones in which the training phase has been performed. This is the biggest limit of RF-based device-free crowd counting and more generally of activity recognition techniques. While it is recognized that overcoming this limit is fundamental, so far very little effort has been spent on such a challenging problem. An approach to overcome this limit has been recently published [11]. The work in [11] is based on the use of CSI and on the statistical analysis of the changes, induced by the presence of people, of the *shape* of the Channel Frequency Response (CFR) calculated on the received signal.

In this paper, we propose and preliminary assess a crowd counting system that is based on the analysis of the Doppler spectrum obtained through the CSI gathered by a single WiFi receiver.

It is worth noting that, even if this work is based on the *Doppler* spectrum, the proposed method is definitely different from a standard or passive bistatic radar approache. In fact, we do not have access to the reference signal [12] and we can not compute the 2D Cross Correlation Function (2D-CCF) between the reference signal (transmitted signal) and the surveillance signal (received signal), achieving the typical range-velocity map. On the other hand, this is not needed. The idea upon which the proposed method is based is that the Doppler spectrum calculated at the receiver changes with the number of people in the room. Intuitively, we expect that the width, the height and the number of peaks of the Doppler spectrum increase as the crowd density increases, as also shown in Fig 1. This intuition will be confirmed by the experimental results. The Doppler spectrum is obtained performing the Fast Fourier Transform (FFT) of the sequence of values of the CSI for each subcarrier and

978-1-5090-0724-0/16/$31.00 ©2016 IEEE

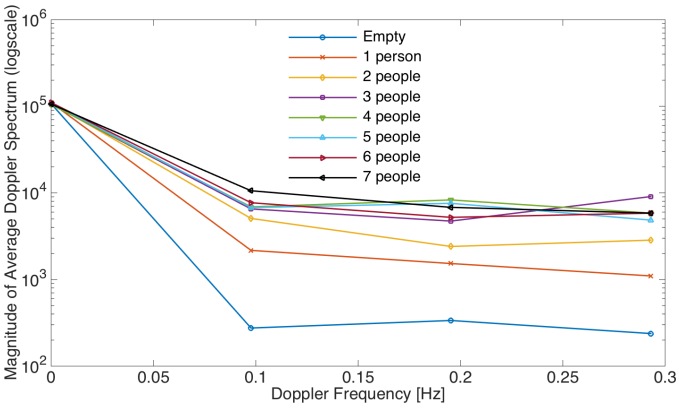

Figure 1: Doppler spectrum obtained for a different number of people in the room.

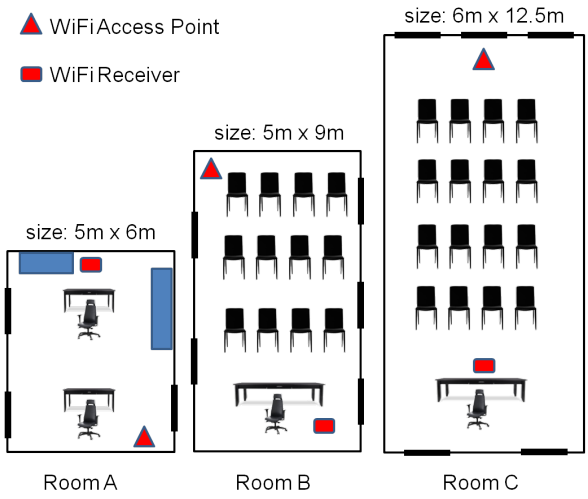

Figure 2: Experimental environments.

then computing the average of the Doppler shift over all the subcarriers. For the classification, we have selected specific shape descriptors of the Doppler spectrum such as decay index or spread.

The results have been achieved under the following test conditions:

- a single WiFi Access Point (AP) and a single WiFi receiver have been used whilst many of the related works use multiple APs/transmitters and receivers;
- volunteers could move as they wanted (randomly, repeating a path, etc.) or they could simply stand;
- the time duration of the rolling window has been set to $W_t = 10$ s that is an acceptable value for the estimation delay. A larger window size allows to increase the accuracy thanks to a better statistical analysis, however, a large window size is helpless when the crowd size changes more frequently than the estimation delay.

In the remainder of this paper, first of all we identify specific features related to the Doppler spectrum obtained from the Doppler-Subcarrier map. Relevant descriptors and antenna combination methods are considered. We then select the features which maximize the classification accuracy. Furthermore, we show, through experimental results, that this method can be used in environments different than the ones over which the training phase has been performed.

II. Architecture and Experimental Setup

In this Section, we present the architecture and the experimental setup of our crowd counting system, which is based on CSI data collected from WiFi AP. The system includes the following modules: a commercial standard 2.4 GHz WiFi b/g/n AP which acts as a double antenna transmitter; a laptop with Ubuntu 10.04 LTS; a WiFi card Intel WiFi Link 5300 with 3 antennas which acts as receiver. A Matlab script running on the same laptop processes offline the collected data and provides the information about the estimated number of people. It is worth noting that the algorithm complexity is quite limited, hence the data processing can be performed in real time on the laptop.

The laptop sends ICMP echo request packets every 20 ms to the WiFi AP and receives the ICMP echo reply packets. CSI are extracted from the received packets by using a customized firmware and an open source Linux wireless driver for the Intel 5300 WiFi card [13], [14].

We carried out WiFi CSI collection on three different rooms, see Figure 2:

- Room A: small size office room (5m $\times$ 6m).
- Room B: medium size meeting room (5m $\times$ 9m).
- Room C: large size meeting room (6m $\times$ 12.5m).

The furniture and the location of the AP and the WiFi receiver are shown in Figure 2. The distance between the AP and the WiFi receiver was larger than 5 m in any room. In order to fairly compare the performance of the system, the number of people during the experiments in the 3 environments under test must be the same. Due to the space constraints of the Room A, which is the smallest one, the crowd size was limited to $N_c = 7$ people. No guidelines or limitations to movements have been given to the volunteers which could walk or stand firm following a random sequence. A number of $N_{CSI} = 5000$ CSI measurements has been collected for each class of people (from 0 to $N_c = 7$) and for each room. The collected CSI data are processed offline to build the dataset, performing data fusion, feature extraction, feature selection and classification tasks.

III. Proposed Approach

In the following Section, we describe the methodology of our CSI-based crowd counting system. First, we present the

source data, i.e. the CSI set, exploited by the crowd counting system. After that, we show which features have been considered and then selected to maximize the performance. Lastly, we describe the classification algorithm.

A. CSI Data

In radio communications, the transmitted electromagnetic waves reach the receiving antenna by more paths. This phenomenon is also known as multipath propagation. The presence and the movements of a human body imprint an information on the intercepted multipath components and hence on the CFR, allowing the recognition of its presence, activity and gesture without wearing any device or sensor.

The proposed crowdcounting system exploits that time variations of the CFR, induced by the presence of people and highlighted by the Doppler Spectrum. The CFR is achieved by the CSI estimated by the WiFi OFDM receiver. An OFDM-based transmission encodes the digital data stream on a set of orthogonal sub-carriers. Let us denote with $\mathbf{y}$ the received complex vector after the FFT at the receiver, $\mathbf{y}$ can be expressed as follows:

$$\mathbf{y} = \mathbf{h_c} \cdot \mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{x}$ is the transmitted complex vector, $\mathbf{h_c}$ is the complex channel response vector in the frequency domain, and $\mathbf{n}$ is the additive Gaussian noise vector. The *i-th* component of $\mathbf{h_c}$ is the CSI value of the i-th OFDM subcarrier, which represents the complex channel gain for the *i-th* subcarrier. To estimate $\mathbf{x}$, the receiver needs to estimate the channel vector $\mathbf{h_c}$, so this estimation is available at the receiver. Therefore, we used the modified firmware of the Intel 5300 Wireless Link card to gather the raw CSI values available at the receiver. This modified firmware provides the CSI value for $N_{sub} = 30$ OFDM subcarriers. The complex vector containing the N_{sub} CSI values, hereinafter CSI vector, can be expressed as:

$$\mathbf{h} = [h_1, h_2, ..., h_{30}], \tag{2}$$

where the CSI of the i-th subcarrier is defined as:

$$h_i = |h_i| e^{j \angle h_i}, \tag{3}$$

where $|h_i|$ is the amplitude and $\angle h_i$ is the phase of the channel of the *i-th* subcarrier. In our experiments we have used 2 transmit antennas and 3 receive antennas, therefore N_{ch} = 6 RF channels are available for processing. This means that, for each received packet, N_{ch} CSI vectors are available to extract the desired features.

B. Feature Extraction

The main idea beyond the proposed crowd-counting system is that an increase of people number is expected to spread the Doppler energy over a larger bandwidth. The Doppler spectrum has been calculated taking into account the CSI vectors collected during the observation interval for N_{pkt} ICMP packets and through all the N_{ch} RF channels, resulting in N_{ch} matrices with N_{sub} rows and N_{pkt} columns:

$$\mathbf{h}^n = \begin{bmatrix} h^n_{1,1} & \dots & h^n_{1,j} & \dots & h^n_{1,N_{pkt}} \\ h^n_{2,1} & \dots & h^n_{2,j} & \dots & h^n_{2,N_{pkt}} \\ \vdots & & \vdots & & \vdots \\ h^n_{N_{sub},1} & \dots & h^n_{N_{sub},j} & \dots & h^n_{N_{sub},N_{pkt}} \end{bmatrix} \tag{4}$$

where $n = 1, \dots, N_{ch}$ and the element $h^n_{i,j}$ corresponds to the CSI extracted from the i-th packet, the j-th subcarrier and the n-th RF channel.

Subsequently, for each row of the matrices $\mathbf{h}^n$, the magnitude of the FFT of the CSI magnitude is computed over a sliding window of size $W = 512$, as follows:

$$H^n_{i,k}(j_0) = \left| \sum_{j=j_0+1}^{j_0+W} |h^n_{i,j}| e^{-i\frac{2\pi}{W}jk} \right| \quad k = 1, \dots, W \tag{5}$$

so as to get a matrix containing the Doppler spectra for each subcarrier:

$$\mathbf{H}^n(j_0) = \begin{bmatrix} H^n_{1,1}(j_0) & \dots & H^n_{1,W}(j_0) \\ H^n_{2,1}(j_0) & \dots & H^n_{2,W}(j_0) \\ \vdots & & \vdots \\ H^n_{N_{sub},1}(j_0) & \dots & H^n_{N_{sub},W}(j_0) \end{bmatrix} \tag{6}$$

where the element $H^n_{i,k}(j_0)$ is the magnitude of the *k*-th spectral component for the *i*-th subcarrier and for the j_0-th sliding window translation.

In order to take into account the contribution of all subcarriers, the N_{sub} spectra have been equally averaged to obtain a mean Doppler spectrum:

$$\begin{aligned} \overline{\mathbf{H}}^n(j_0) &= \frac{1}{N_{sub}} \sum_{i=1}^{N_{sub}} [H^n_{i,1}(j_0), ..., H^n_{i,k}(j_0), ..., H^n_{i,W}(j_0)] = \\ &= [H^n_1(j_0), ..., H^n_k(j_0), ..., H^n_W(j_0)] \end{aligned} \tag{7}$$

where $H^n_k(j_0)$ is the mean *k*-th Doppler spectrum component for the *n*-th channel. In the following notation we will not explicit the dependence from j_0.

Once the mean Doppler spectrum $\overline{\mathbf{H}}^n(j_0)$ has been calculated, its shape must be characterized and, hence, the *spectral descriptors*, indicated in Table I, for the n-th RF channel have been considered.

Since the descriptors D have been calculated for the N_{ch} channels they have to be properly combined through a function $g(\cdot)$ in order to get *spectral features* F taking into account all RF channels.

Descriptor	Formula	Description
Mean	$\mu_n = \frac{1}{W} \sum_{k=1}^{W} H_k^n$	The arithmetic mean of the doppler spectrum.
Standard Deviation	$\sigma_n = \frac{1}{W-1} \sum_{k=1}^{W} (H_k^n - \mu_n)^2$	The standard deviation of the doppler spectrum.
Spectral Centroid	$f_k = k\frac{25}{W}$ Hz, $\quad k = 1, \ldots, W$ $\mathrm{SC}_n = \frac{\sum_{k=1}^{W} [H_k^n]^2 f_k}{\sum_{k=1}^{W} [H_k^n]^2}$	The spectral centroid indicates the position of the "center of mass" of the Doppler spectrum. It is calculated as the weighted mean of the frequencies present in the signal with their magnitudes as the weights.
Decay Factor	$\alpha_n = \sum_{k=1}^{W} f_k H_k^n$ $\lambda_n = -\frac{1}{W-1} \sum_{k=2}^{W} \frac{H_k^n - H_{k-1}^n}{f_k - f_{k-1}} \frac{2}{H_k^n + H_{k-1}^n}$	The Decay parameter α represents the area under each doppler spectrum curve, while the parameter λ is calculated as the mean of the derivative function) for each doppler spectrum curve.
Spectral Entropy	$E_{tot} = \sum_{k=1}^{W} [H_k^n]^2$ $E_{t-blk} = \sum_{k=(t-1)S+1}^{tS} [H_k^n]^2$ $\mathrm{SE}_n = -\sum_{t=1}^{N_{blk}} \frac{E_{t-blk}}{E_{tot}} \log_2 \left[\frac{E_{t-blk}}{E_{tot}}\right]$	The Spectral Entropy represents the amount of information contained in the doppler spectrum.
Spectral Flatness	$\mathrm{SF}_n = \frac{\sqrt[W]{\prod_{k=1}^{W} [H_k^n]^2}}{\frac{1}{W} \sum_{k=1}^{W} [H_k^n]^2}$	The Spectral Flatness is a measure used in digital signal processing to quantify how noise-like a signal is.
Spectral Rolloff	$\mathrm{SR}_n = \hat{k} : \sum_{k=1}^{\hat{k}} [H_k^n]^2 \geq 0.9 E_{tot}$	The Spectral Rolloff is the frequency below which the 90% of the accumulated magnitude of the spectrum is concentrated.
Spectral Slope	$\overline{f}_k = \frac{1}{W} \sum_{k=1}^{W} f_k$ $\mathrm{SSL}_n = \frac{\sum_{k=1}^{W} (f_k - \overline{f}_k)(H_k^n - \overline{H}_k^n)}{\sum_{k=1}^{W} (f_k - \overline{f}_k)^2}$	The Spectral Slope is a measure of the slope of the spectral shape.
Spectral Spread	$\mathrm{SSP}_n = \sqrt{\frac{\sum_{k=1}^{W} [H_k^n]^2 (f_k - \mathrm{SC}_n)^2}{\sum_{k=1}^{W} [H_k^n]^2}}$	The Spectral Spread is defined as the second central moment of the log-frequency spectrum and gives indications about how the spectrum is distributed around its centroid.

2-nd Order Spectral Moment	$\eta_{2,n} = \frac{\sum_{k=1}^{W} [H_k^n]^2 f_k^2}{\sum_{k=1}^{W} [H_k^n]^2}$	The 2-nd order spectral moment of the doppler spectrum curve.
2-nd Order Spectral Central Moment	$\xi_{2,n} = \frac{\sum_{k=1}^{W} [H_k^n]^2 (f_k - \mathrm{SC}_n)^2}{\sum_{k=1}^{W} [H_k^n]^2}$	The 2-nd order spectral central moment of the doppler spectrum curve.
3-rd Order Spectral Moment	$\eta_{3,n} = \frac{\sum_{k=1}^{W} [H_k^n]^2 f_k^3}{\sum_{k=1}^{W} [H_k^n]^2}$	The 3-rd order spectral moment of the doppler spectrum curve.
3-rd Order Spectral Central Moment	$\xi_{3,n} = \frac{\sum_{k=1}^{W} [H_k^n]^2 (f_k - \mathrm{SC}_n)^3}{\sum_{k=1}^{W} [H_k^n]^2}$	The 3-rd order spectral central moment of the doppler spectrum curve.
4-th Order Spectral Moment	$\eta_{4,n} = \frac{\sum_{k=1}^{W} [H_k^n]^2 f_k^4}{\sum_{k=1}^{W} [H_k^n]^2}$	The 4-th order spectral moment of the doppler spectrum curve.
4-th Order Spectral Central Moment	$\xi_{4,n} = \frac{\sum_{k=1}^{W} [H_k^n]^2 (f_k - \mathrm{SC}_n)^4}{\sum_{k=1}^{W} [H_k^n]^2}$	The 4-th order spectral central moment of the doppler spectrum curve.
Spectral Skewness	$T_{k,n} = \frac{f_k - SC_n}{\sqrt{\xi_{2,n}}}, \qquad k = 1, \ldots, W$ $\mathrm{Skew}_n = \frac{\sum_{k=1}^{W} [H_k^n]^2 T_{k,n}^3}{\sum_{k=1}^{W} [H_k^n]^2}$	The Spectral Skewness is a measure of the asymmetry of the doppler spectrum about its mean.
Spectral Kurtosis	$\mathrm{Kurt}_n = \frac{\sum_{k=1}^{W} [H_k^n]^2 T_{k,n}^4}{\sum_{k=1}^{W} [H_k^n]^2}$	The Spectral Kurtosis is a measure of the "tailedness" of the doppler spectrum.
Mean-Sigma Ratio	$\rho_n = \frac{\mu_n}{\sigma_n}$	The Mean-Sigma Ratio represents the ratio between the arithmetic mean of the doppler spectrum and its standard deviation.

Table I: List of descriptors.

$$F = g(D_1, D_2, \ldots, D_{N_{ch}}) = \sum_{n=1}^{N_{ch}} D_n w_n \qquad (8)$$

In particular, several combination functions have been considered, which also exploit the RSSI trend relative to the single channels and are indicated in Table II.

C. Feature Selection and Classification

In our approach we used an exhaustive search feature selection strategy to reduce the number of features and to improve the overall classification accuracy, which were taken as objective function for the selection task. Starting from the previously mentioned set of features, we have computed the mean classification accuracy over the test environments for each feature and for each possible combinations of them. Then, we have chosen the features subset which is the best trade-off between the overall recognition accuracy and the feature space dimension. As a result, the selected feature is composed as follows:

Combining scheme	Formula
Arithmetic Mean	$w_n = \frac{1}{N_{ch}}$
RSSI Standard Deviation	$w_n = \frac{\sigma_{\text{RSSI},n}}{\sum_{l=1}^{N_{ch}} \sigma_{\text{RSSI},l}}$
RSSI Variance	$w_n = \frac{\sigma^2_{\text{RSSI},n}}{\sum_{l=1}^{N_{ch}} \sigma^2_{\text{RSSI},l}}$
RSSI Third-Power Standard Deviation	$w_n = \frac{\sigma^3_{\text{RSSI},n}}{\sum_{l=1}^{N_{ch}} \sigma^3_{\text{RSSI},l}}$
RSSI Fourth-Power Standard Deviation	$w_n = \frac{\sigma^4_{\text{RSSI},n}}{\sum_{l=1}^{N_{ch}} \sigma^4_{\text{RSSI},l}}$
RSSI Standard Deviation over Mean	$w_n = \frac{\frac{\sigma_{\text{RSSI},n}}{\mu_{\text{RSSI},n}}}{\left(\sum_{l=1}^{N_{ch}} \frac{\sigma_{\text{RSSI},l}}{\mu_{\text{RSSI},l}}\right)}$
RSSI Fano Factor	$w_n = \frac{\frac{\sigma^2_{\text{RSSI},n}}{\mu_{\text{RSSI},n}}}{\left(\sum_{l=1}^{N_{ch}} \frac{\sigma^2_{\text{RSSI},l}}{\mu_{\text{RSSI},l}}\right)}$

Table II: List of combining schemes.

$$F = \frac{1}{N_{ch}} \sum_{n=1}^{N_{ch}} \text{Kurt}_n \tag{9}$$

The exhaustive search has also shown that increasing the number of features does not lead to a significant improvement of the classification accuracy. This means that using more features does not introduce more useful information to classify different human crowd sizes. In other words, the shape-related features are quite correlated with each other, hence the employment of a large feature set does not significantly improve the system performance. Experimental results show that using 3 features instead of 1 increases the mean overall accuracy of less 2%, while using 4 features the accuracy improvement is of about 4%.

Since our effort has been focused on the feature extraction phase, for the estimation of the crowd density and of the number of people we have chosen the Naive Bayes classifier, which is a simple probabilistic classifier. The Naive Bayes classifier is based on the application of the Bayes' theorem and relies on the independence among features. Each feature distribution is well approximated by a normal one, hence, this kind of distribution has been used to build the classifier. During the training step for each class the parameters of the Gaussian distribution, i.e. mean and standard deviation, also called *likelihood*, are estimated using the training data. Afterwards, in the test step the algorithm classifies any new sample maximizing the posterior probability of that sample belonging to each class.

IV. Experimental Results

The classifier was trained using the full dataset collected in Room B, in fact it represents the middle size room with respect to Room A and Room C. Classification tests in Room A and Room C have been performed using the overall datasets of the corresponding room. Classification results are presented in Table III and IV in terms of confusion matrix for the crowd density estimation case.

Truth \ Prediction	Empty	1 person	2 people	3,4 people	5,6,7 people
Empty	1	0	0	0	0
1 person	0	0.75	0.25	0	0
2 people	0	0.05	0.61	0.34	0
3,4 people	0	0	0.06	0.66	0.28
5,6,7 people	0	0	0	0.38	0.62

0.73

Table III: Confusion matrix for the Room A

Truth \ Prediction	Empty	1 person	2 people	3,4 people	5,6,7 people
Empty	1	0	0	0	0
1 person	0.10	0.90	0	0	0
2 people	0	0.72	0.24	0.04	0
3,4 people	0	0.14	0.34	0.51	0.01
5,6,7 people	0	0.02	0.07	0.43	0.48

0.63

Table IV: Confusion matrix for the Room C

Good and comparable performance in terms of average accuracy are achieved both in Room A and Room C, which respectively are 73% and 63%. This is in line with the fact that the accuracy of Room A is higher than that of Room C even in the case of dedicated training. In fact, we have also noticed that if an environment has better performance than another one in the case of dedicated training, then that environment will also have an higher average accuracy in the trained-once case. Moreover, we found that performing a dedicated training phase for each room, the respective accuracies are higher of about 10% than those obtained with the trained-once approach. It is worth noting that for both rooms, the empty and one person scenarios are recognized by the system with high accuracy. This result is due to the fact that one person walking in the environment highly spreads the Doppler spectrum curve with respect to the empty scenario. In fact, in the latter case no persons are moving, and hence the frequency subcarriers are not affected by any Doppler shifts except by the time variations caused by the noise. This means that the two considered scenarios

have a very different Doppler curve and then they are easily recognizable by the selected shape features.

Comparing the confusion matrices for Room A and Room C, we can notice that the recognition accuracy for the single classes remains quite constant moving from 1 person to 5,6,7 persons in Room A, while this is not true in Room C where the recognition accuracy decreases rapidly moving from 1 person to 5,6,7 persons. This is also in line with the results achieved in [11].

For what concerns the crowd counting case, the results show that more than 84% and 71% estimation errors are less or equal than 1 person in Room A and Room C respectively. Moreover, the errors are less or equal than 2 persons for 99% and 92% of the time in Room A and Room C, respectively.

V. Related Works and Estimation Accuracies

This Section provides an overview of the most important related works. They can be divided into two main categories: RSSI-based and CSI-based. As a matter of fact, most of the works are based on RSSI measurements.

Depatla et al. developed a mathematical expression for the probability distribution of the received signal amplitude as a function of the total number of occupants, which is the base for people counting using Kullback-Leibler divergence [5]. A single AP and a single receiver are used and the estimation delay is 300 s. Two sites are used in the experiments: an indoor site with the dimension 4.4×7.5 m^2 and an outdoor site with the dimension 7×10 m^2. Experiments performed with up to 9 people show that an error of 2 or less is achieved in 96% and 63% of the time for the outdoor and indoor cases respectively, when using the typical omnidirectional antennas that come as part of the standard WiFi cards.

Xu et al. deployed an infrastructure of 20 to 22 devices to count and localize subjects in large areas [6]. The radio devices used in the experiments contain a Chipcon CC1100 radio transceiver operating in the unlicensed band at 909.1 MHz. Each transmitter periodically broadcasts a 10-byte packet every 100 ms. When the receiver receives a packet, it measures the RSS values and wraps the transmitter ID, receiver ID, RSS, timestamp (on the receiver side) into a service packet which is sent to a centralized system for data collection and analysis. Experimental results show that SCPL works well in two different typical indoor environments of 150 m^2 (office cubicles) and 400 m^2 (open floor plan). In both spaces, we can achieve about an 86% accuracy for up to 4 subjects.

Yuan et al. deployed a grid of 16 sensor nodes (TelosB) in a room with dimension 18×18 m^2 [7]. A database server collects the signal strengths and uses the information to cluster the crowd density level in the room. An average accuracy of 86% was achieved to estimate three types of densities are estimated: low (0-3 persons), moderate (4-10 persons) and high ($>$10 persons).

Fadhlullah et al. deployed three ZigBee wireless nodes continuously reporting the received signal strength indicator to the main node [8]. The average RSSI was calculated using the recorded information of all tags and classified into three crowd density categories which are low (0-5 persons), medium (6-10 persons), and high (11-15 persons). The results showed that the system was 75% and 70% accurate in detecting the low and medium human crowd density, respectively.

Yoshida et al. applied a regression-based approach to count the number of people using the RSSI collected by 10 nodes using one AP in a single room [9]. The estimation delay is 300 s. The accuracy of support vector regression-based method for estimating the number of people is 77% for up to 7 people.

In [10] authors propose a device-free flow counting system based on RSSI measurements in a wireless sensor network. The experiment has been performed only in a single indoor environment using one HBE-ZIGBEX mote as transmitter and two HBE-ZIGBEX motes as receivers. The number of people was limited to 5 and the walking paths were predefined. In the best case the mean prediction accuracy was 77% with a standard deviation of about 1%.

The CSI-based crowd counting method proposed in literature are [11] and [13]. In the latter on, authors applied a Gray Verhulst model to construct a normal profile of the percentage of nonzero elements of the dilated matrix of CSIs for each class. The experiments are carried out counting up to 30 persons in both indoor and outdoor settings using a single AP and 3 or 4 receivers. The results show that this approach achieves an error ranging from 2 persons to three persons; more than 98% estimation errors are less than 2 persons in indoor environment, while about 70% errors is 2 persons in outdoor environment.

Comparison of performance and experimental setup among crowd counting methods including our method is shown in Table V. It is worth noting that all the reported methods, except our method, perform a training phase in the same room where they perform the crowd counting test.

VI. Conclusion

This paper presents a WiFi-based device-free crowd counting system that can be used in rooms/environments different than those in which the training has been performed and without requiring another training phase. The proposed approach is based on the intuition that regardless the background environment, the peaks of the Doppler spectrum increase as the crowd density increases. The experimental results in terms of confusion matrices prove the effectiveness of the proposed approach for crowd density and crowd counting estimation. In particular, in case of crowd density estimation, average accuracies are 73% and 63% in the two different environments.

Table V: Comparison of device-free crowd counting methods

Ref	Environments	Max # of People	Standards	Source Data	# TXs	# RXs	Estimation Delay (window size W)	Crowd Counting Accuracy	Crowd Density Accuracy
[5]	2 (indoor 33 m^2, outdoor 70 m^2)	9	WiFi	RSSI	1	1	300 s	P(e$\leq$ 2) = 96%(outdoor) P(e$\leq$ 2) = 63%(indoor)	-
[6]	2 (indoor 150 m^2, indoor 400 m^2)	4	ISM 909.1 MHz	RSSI	12/13	8/9	1 s	84%	-
[7]	1 (indoor 324 m^2)	>10	ISM 2.4 GHz	RSSI	16	16	-	-	86% (0p-3p, 4p-10p, ¿10p)
[6]	2 (indoor 150 m^2, indoor 400 m^2)	4	ISM 909.1 MHz	RSSI	12/13	8/9	1 s	84%	-
[7]	1 (indoor 324 m^2)	>10	ISM 2.4 GHz	RSSI	16	16	-	-	86% (0p-3p, 4p-10p, ¿10p)
[8]	1 (indoor 100 m^2)	15	ZigBee 2.4 GHz	RSSI	1	3	-	-	73% (5p, 10p, 15p)
[9]	1 (indoor)	7	WiFi	RSSI	1	10	300 s	77%	94% (0p, 1p-3p, 4p-7p)
[10]	1 (indoor corridor $\approx$ 10 m^2)	5	HBE-Zigbex	RSSI	1	2	4.5 s	77%	-
[13]	2 (indoor, outdoor)	30	WiFi	CSI	1	3/4	-	P(e$\leq$ 2) = 70% (outdoor) P(e$\leq$ 2) = 98% (indoor)	-
Our system	3 (indoor 30 m^2, indoor 45 m^2, indoor 70 m^2)	7	WiFi	CSI	1	1	10 s	P(e$\leq$2) = 99% (room A) P(e$\leq$2) = 92%(room C)	73% (room A) 63% (room C) (0p, 1p, 2p, 3p-4p, 5p-7p)

For the crowd counting case, instead, the probability that estimation errors are less or equal to 1 person are 84% and 71% for Room A and Room C respectively. Moreover, the probabilities that the errors are less or equal to 2 persons are 99% and 92%.

REFERENCES

[1] Y. Wang, J. Yang, Y. Chen, H. Liu, M. Gruteser, and R. P. Martin, "Tracking human queues using single-point signal monitoring," in *Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '14, 2014, pp. 42–54.

[2] I. Bisio, F. Lavagetto, M. Marchese, and A. Sciarrone, "Smartphone-centric ambient assisted living platform for patients suffering from co-morbidities monitoring," *Communications Magazine, IEEE*, vol. 53, no. 1, pp. 34–41, January 2015.

[3] M. De Sanctis, E. Cianca, S. Di Domenico, D. Provenziani, G. Bianchi, and M. Ruggieri, "Wibecam: Device free human activity recognition through wifi beacon-enabled camera," in *Proceedings of the 2nd Workshop on Physical Analytics*, ser. WPA '15, 2015, pp. 7–12.

[4] L. Militano, M. Condoluci, G. Araniti, A. Molinaro, A. Iera, and G. M. Muntean, "Single frequency-based device-to-device-enhanced video delivery for evolved multimedia broadcast and multicast services," *IEEE Transactions on Broadcasting*, vol. 61, no. 2, pp. 263–278, June 2015.

[5] S. Depatla, A. Muralidharan, and Y. Mostofi, "Occupancy estimation using only wifi power measurements," *Selected Areas in Communications, IEEE Journal on*, vol. 33, no. 7, pp. 1381–1393, July 2015.

[6] C. Xu, B. Firner, R. S. Moore, Y. Zhang, W. Trappe, R. Howard, F. Zhang, and N. An, "Scpl: Indoor device-free multi-subject counting and localization using radio signal strength," in *Proceedings of the 12th International Conference on Information Processing in Sensor Networks*. New York, NY, USA: ACM, 2013, pp. 79–90. [Online]. Available:

[7] Y. Yuan, J. Zhao, C. Qiu, and W. Xi, "Estimating crowd density in an rf-based dynamic environment," *Sensors Journal, IEEE*, vol. 13, no. 10, pp. 3837–3845, Oct 2013.

[8] S. Y. Fadhlullah and W. Ismail, "A statistical approach in designing an rf-based human crowd density estimation system," *International Journal of Distributed Sensor Networks*, vol. 2016, pp. 8 351 017:1–8 351 017:9, 2016. [Online]. Available:

[9] T. Yoshida and Y. Taniguchi, "Estimating the number of people using existing wifi access point in indoor environment," in *Proceedings of the 6th European Conference of Computer Science (ECCS '15)*, 2015, pp. 46–53.

[10] S. H. Doong, "Spectral human flow counting with rssi in wireless sensor networks," in *2016 International Conference on Distributed Computing in Sensor Systems (DCOSS)*, May 2016, pp. 110–112.

[11] S. Di Domenico, M. De Sanctis, E. Cianca, and G. Bianchi, "A trained-once crowd counting method using differential wifi channel state information," in *Proceedings of the 3rd International on Workshop on Physical Analytics*, ser. WPA '16, 2016, pp. 37–42.

[12] F. Colone, P. Falcone, C. Bongioanni, and P. Lombardo, "Wifi-based passive bistatic radar: Data processing schemes and experimental results," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 48, no. 2, pp. 1061–1079, APRIL 2012.

[13] W. Xi, J. Zhao, X. Li, K. Zhao, S. Tang, X. Liu, and Z. Jiang, "Electronic frog eye: Counting crowd using wifi," in *2014 IEEE Conference on Computer Communications, INFOCOM 2014, Toronto, Canada*, 2014, pp. 361–369.

[14] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Predictable 802.11 packet delivery from wireless channel measurements." in *Proceedings of the Second International Workshop on Wireless Network Measurement*, ser. CM SIGCOMM 2010, 2010, pp. 159–170.